

# Stochastic Equicontinuity in Nonlinear Time Series Models

Andreas Hagemann\*

*University of Illinois at Urbana-Champaign*

March 8, 2013

In this note I provide simple and easily verifiable conditions under which a strong form of stochastic equicontinuity holds in a wide variety of modern time series models. In contrast to most results currently available in the literature, my methods avoid mixing conditions. I discuss two applications in detail.

## 1. Introduction

Stochastic equicontinuity typically captures the key difficulty in weak convergence proofs of estimators with non-differentiable objective functions. Precise and elegant methods have been found to deal with cases where the data dependence structure can be described by mixing conditions; see Dedecker et al. (2007) for an excellent summary. Mixing assumptions are convenient in this context because they measure how events generated by time series observations—rather than the observations themselves—relate to one another and therefore also measure dependence of functions of such time series. The downside to these assumptions is that they can be hard to verify for a given application. Hansen (1996) describes alternatives and considers parametric classes of functions that behave like mixingales, but his results come at the expense of Lipschitz continuity conditions on these functions and rule out many applications of interest.

In this note I give simple and easily verifiable conditions under which objective functions of econometric estimators are stochastically equicontinuous when the underlying process is a stationary time series of the form

$$\xi_i = \xi(\varepsilon_i, \varepsilon_{i-1}, \varepsilon_{i-2}, \dots). \quad (1.1)$$

---

\*Department of Economics, University of Illinois at Urbana-Champaign, 214 David Kinley Hall, 1407 W. Gregory, Urbana, Illinois 61801, MC-707, USA. Tel.: +1-217-333-0120. Fax: +1-217-244-6571. E-mail address: [ahagama2@illinois.edu](mailto:ahagama2@illinois.edu). I would like to thank Roger Koenker for his continued support and encouragement. I gratefully acknowledge financial support from a UIUC Summer Research Fellowship. All errors are my own.

Here  $(\varepsilon_i)_{i \in \mathbb{Z}}$  is a sequence of iid copies of a random variable  $\varepsilon$  and  $\xi$  is a measurable, possibly unknown function that transforms the input  $(\varepsilon_i, \varepsilon_{i-1}, \dots)$  into the output  $\xi_i$ . The stochastic equicontinuity problem does not have to be parametric and no continuity conditions are needed. The class (1.1) allows for the construction of dependence measures that are directly related to the stochastic process and includes a large number of commonly-used stationary time series models. The next section provides several specific examples.

In the following,  $\|X\|_p$  denotes  $(\mathbb{E}|X|^p)^{1/p}$  and  $\mathbb{P}^*$  and  $\mathbb{E}^*$  are outer probability and expectation, respectively (see van der Vaart, 1998, p. 258). Limits are as  $n \rightarrow \infty$ .

## 2. Stochastic Equicontinuity in Nonlinear Time Series Models

Let  $\nu_n f := n^{-1/2} \sum_{i=1}^n (f(\xi_i) - \mathbb{E}f(\xi_0))$  be the empirical process evaluated at some function  $f$ . Here  $f$  is a member of a class of real-valued functions  $\mathcal{F}$ . In econometric applications,  $\mathcal{F}$  is typically a parametric class  $\{f_\theta : \theta \in \Theta\}$ , where  $\Theta$  is a bounded subset of  $\mathbb{R}^k$ , although no parametric restriction on  $\mathcal{F}$  is necessary in the following. Define a norm by  $\rho(f) = \|f(\xi_0)\|_2$ . An empirical process is said to be *stochastically equicontinuous* (see, e.g., Pollard, 1985, p. 139) on  $\mathcal{F}$  if for all  $\epsilon > 0$  and  $\eta > 0$ , there is a  $\delta > 0$  such that

$$\limsup_{n \rightarrow \infty} \mathbb{P}^* \left( \sup_{f, g \in \mathcal{F}: \rho(f-g) < \delta} |\nu_n(f-g)| > \eta \right) < \epsilon. \quad (2.1)$$

As mentioned above, proving stochastic equicontinuity is often the key difficulty in weak convergence proofs. The next two examples illustrate typical applications.

**Example 1** (Quantilograms). Linton and Whang (2007) measure the directional predictive ability of stationary time series  $(X_i)_{i \in \mathbb{Z}}$  with the quantilogram, a normalized version of  $\mathbb{E}(\alpha - 1\{X_0 < \theta_\alpha\})(\alpha - 1\{X_h < \theta_\alpha\})$  with  $\alpha \in (0, 1)$  and  $h = 1, 2, \dots$ , where  $\theta_\alpha$  is the  $\alpha$ -quantile of the marginal distribution of  $(X_i)_{i \in \mathbb{Z}}$ . Let  $\xi_i = (X_{i-h}, X_i)$  and  $f_\theta(\xi_i) = (\alpha - 1\{X_{i-h} < \theta\})(\alpha - 1\{X_i < \theta\})$ . Under the null hypothesis of no directional predictability, we have  $\mathbb{E}f_{\theta_\alpha}(\xi_0) = 0$  for all  $h = 1, 2, \dots$ . Let  $\hat{\theta}_{n,\alpha}$  be the sample  $\alpha$ -quantile and replace population moments by sample moments to obtain  $(n-h)^{-1} \sum_{i=1+h}^n f_{\hat{\theta}_{n,\alpha}}(\xi_i)$ , the sample version of  $\mathbb{E}f_{\theta_\alpha}(\xi_0)$ . Apart from a scaling factor, the asymptotic null distribution of the sample quantilogram can be determined through the decomposition

$$(n-h)^{-1/2} \sum_{i=1+h}^n f_{\hat{\theta}_{n,\alpha}}(\xi_i) = \sqrt{n-h} \mathbb{E}f_{\hat{\theta}_{n,\alpha}}(\xi_0) + \nu_{n-h} f_{\theta_\alpha} + \nu_{n-h}(f_{\hat{\theta}_{n,\alpha}} - f_{\theta_\alpha}).$$

If the distribution of  $X_i$  is smooth, the delta method can be used to control the first term on the right and, under dependence conditions, an ordinary central limit theorem applies to the second term. Further, we have  $\rho(f_{\hat{\theta}_{n,\alpha}} - f_{\theta_\alpha}) \rightarrow_p 0$  whenever  $\hat{\theta}_{n,\alpha} \rightarrow_p \theta_\alpha$  (see Example 3 below). Hence, we can take  $\mathcal{F} = \{f_\theta : \theta \in \Theta\}$ , where  $\Theta$  is a compact neighborhood of  $\theta_\alpha$ , and as long as (2.1) holds, the third term on the right-hand side of the preceding display

converges to zero in probability because in large samples

$$\mathbb{P}\left(\nu_{n-h}(f_{\hat{\theta}_{n,\alpha}} - f_{\theta_\alpha}) > \eta, \rho(f_{\hat{\theta}_{n,\alpha}} - f_{\theta_\alpha}) < \delta\right) \leq \mathbb{P}^*\left(\sup_{f_\theta \in \mathcal{F}: \rho(f_\theta - f_{\theta_\alpha}) < \delta} |\nu_{n-h}(f_\theta - f_{\theta_\alpha})| > \eta\right).$$

**Example 2** (Robust  $M$ -estimators of location). Robust location estimators can often be defined implicitly as an  $M$ -estimator  $\hat{\theta}_n$  that nearly solves  $n^{-1} \sum_{i=1}^n f_\theta(\xi_i) = 0$  in the sense that  $\sum_{i=1}^n f_{\hat{\theta}_n}(\xi_i) = o_p(\sqrt{n})$ . Popular examples include the median with  $f_\theta(x) = \text{sign}(x - \theta)$  and Huber estimators with  $f_\theta(x) = -\Delta 1\{x - \theta < -\Delta\} + (x - \theta)1\{|x - \theta| \leq \Delta\} + \Delta 1\{x - \theta > \Delta\}$  for some  $\Delta > 0$ . Add and subtract in  $\sum_{i=1}^n f_{\hat{\theta}_n}(\xi_i) = o_p(\sqrt{n})$  to see that stochastic equicontinuity implies  $\sqrt{n} \mathbb{E} f_{\hat{\theta}_n}(\xi_0) + \nu_n f_{\theta_0} = o_p(1)$ . The limiting behavior of  $\sqrt{n}(\hat{\theta}_n - \theta_0)$  can then again be determined through the delta method and a central limit theorem.

Stochastic equicontinuity cannot hold without restrictions on the complexity of the set  $\mathcal{F}$ ; see, e.g., Andrews (1994, pp. 2252–2253). Here, complexity of  $\mathcal{F}$  is measured via its bracketing number  $N = N(\delta, \mathcal{F})$ , the smallest number for which there are functions  $f_1, \dots, f_N \in \mathcal{F}$  and functions  $b_1, \dots, b_N$  (not necessarily in  $\mathcal{F}$ ) such that  $\rho(b_k) \leq \delta$  and  $|f - f_k| \leq b_k$  for all  $1 \leq k \leq N$ . In addition, some restrictions are required on the memory of the time series. For processes of the form (1.1), the memory is most easily controlled by comparing  $\xi_i$  to a slightly perturbed version of itself (see Wu, 2005). Let  $(\varepsilon_i^*)_{i \in \mathbb{Z}}$  be an iid copy of  $(\varepsilon_i)_{i \in \mathbb{Z}}$ , so that the difference between  $\xi_i$  and  $\xi'_i := \xi(\varepsilon_i, \dots, \varepsilon_1, \varepsilon_0^*, \varepsilon_{-1}^*, \dots)$  are the inputs prior to period 1. Assume the following:

**Assumption A.** *Let  $\mathcal{F}$  be a uniformly bounded class of real-valued functions with bracketing numbers  $N(\delta, \mathcal{F}) < \infty$ . Then there exists some  $\alpha \in (0, 1)$  and  $p > 0$  such that*

- (i)  $\sup_{f \in \mathcal{F}} \|f(\xi_n) - f(\xi'_n)\|_p = O(\alpha^n)$  and
- (ii)  $\max_{1 \leq k \leq N(\delta, \mathcal{F})} \|b_k(\xi_n) - b_k(\xi'_n)\|_p = O(\alpha^n)$  for any given  $\delta > 0$ .

*Remarks.* (i) The examples at the end of this section show that Assumption A often represents only a mild restriction on the dependence structure.

(ii) Because  $\mathcal{F}$  is assumed to be uniformly bounded, the bounding functions  $b_k$  can be chosen to be bounded as well. Hence, in view of Lemma 2 of Wu and Min (2005), the exact choice of  $p$  in Assumption A is irrelevant, for if the assumption holds for some  $p$ , then it holds for all  $p > 0$ .

Assumption A and a complexity requirement on  $\mathcal{F}$  given by a bracketing integral imply a strong form of stochastic equicontinuity. The following theorem (the proof of which is found in the Appendix) is similar to Andrews and Pollard's (1994) Theorem 2.2 with their mixing condition replaced by Assumption A. It implies (2.1) via the Markov inequality.

**Theorem.** *Suppose that Assumption A holds and  $\int_0^1 x^{-\gamma/(2+\gamma)} N(x, \mathcal{F})^{1/Q} dx < \infty$  for some  $\gamma > 0$  and an even integer  $Q \geq 2$ . Then for every  $\epsilon > 0$ , there is a  $\delta > 0$  such that*

$$\limsup_{n \rightarrow \infty} \mathbb{E}^* \left( \sup_{f, g \in \mathcal{F}: \rho(f-g) < \delta} |\nu_n(f - g)| \right)^Q < \epsilon.$$

A useful feature of this theorem is that the constants  $\gamma$  and  $Q$  are not connected to the dependence measures as in Andrews and Pollard (1994). In contrast to their result for mixing arrays,  $\gamma$  and  $Q$  can therefore be chosen to be as small and large, respectively, as desired to make the bracketing integral converge without restricting the set of time series under consideration.

Before concluding this note, the next two examples illustrate how to apply the theorem and how to verify Assumption A in practice.

**Example 3** (Quantilograms, continued). Suppose for simplicity that  $F_X(\theta) := \mathbb{P}(X_0 \leq \theta)$  is Lipschitz on  $\Theta$ . Take a grid of points  $\min \Theta := \theta_0 < \theta_1 < \dots < \theta_N =: \max \Theta$  and let  $b_k(\xi_i) = 1\{X_{i-h} < \theta_k\} - 1\{X_{i-h} < \theta_{k-1}\} + 1\{X_i < \theta_k\} - 1\{X_i < \theta_{k-1}\}$ . Given a  $\theta \in \Theta$ , we can then find an index  $k$  such that  $|f_\theta - f_{\theta_k}| \leq b_k$ , where I used the fact that  $|\alpha - 1\{\cdot\}| \leq \max\{\alpha, 1 - \alpha\} < 1$ . Moreover, by stationarity

$$\rho(b_k) \leq 2\|1\{X_0 < \theta_k\} - 1\{X_0 < \theta_{k-1}\}\|_2 \leq 2\sqrt{F_X(\theta_k) - F_X(\theta_{k-1})},$$

which is bounded above by a constant multiple of  $\sqrt{\theta_k - \theta_{k-1}}$  due to Lipschitz continuity. Hence, if  $\rho(b_k) \leq \delta$  for all  $k = 1, \dots, N$ , we can choose bracketing numbers with respect to  $\rho$  of order  $N(\delta, \mathcal{F}) = O(\delta^{-2})$  as  $\delta \rightarrow 0$  (see Andrews and Pollard, 1994; van der Vaart, 1998, pp. 270–272) and the bracketing integral converges, e.g., for  $\gamma = 1$  and  $Q = 4$ . By the same calculations as in the preceding display, all  $\theta, \theta' \in \Theta$  satisfy  $\rho(f_\theta - f_{\theta'}) = O(|\theta - \theta'|^{1/2})$  as  $\theta \rightarrow \theta'$  and therefore  $\rho(f_{\hat{\theta}_{n,\alpha}} - f_{\theta_\alpha}) \rightarrow_p 0$  if  $\hat{\theta}_{n,\alpha} \rightarrow_p \theta_\alpha$ . In addition, suppose that the geometric contraction (GMC) property of Wu and Min (2005) holds, i.e., there is some  $\beta \in (0, 1)$  and  $p > 0$  such that  $\|\xi_n - \xi'_n\|_p = O(\beta^n)$ . Then Assumption A(i) is also satisfied because  $\|f_\theta(\xi_n) - f_\theta(\xi'_n)\|_p \leq 2\|1\{X_n < \theta\} - 1\{X'_n < \theta\}\|_p + 2\|1\{X_{n-h} < \theta\} - 1\{X'_{n-h} < \theta\}\|_p = O(\alpha^n)$  uniformly in  $\theta$  for some  $\alpha \in (0, 1)$  by Proposition 3.1 of Hagemann (2011). The GMC property holds, e.g., for stationary (causal) ARMA, ARCH, GARCH, ARMA-ARCH, ARMA-GARCH, asymmetric GARCH, generalized random coefficient autoregressive, and quantile autoregressive models; see Shao and Wu (2007) and Shao (2011) for proofs and more examples. All of these models therefore also satisfy Assumption A(i). The same reasoning applies to  $b_k$ .

**Example 4** (Robust  $M$ -estimators of location, continued). Nearly identical arguments as in the preceding example yield stochastic equicontinuity for the median. For the Huber estimator, take the grid from before and note that we can find a  $k$  such that  $|f_\theta - f_{\theta_k}| \leq \min\{\theta_k - \theta_{k-1}, 2\Delta\} =: b_k$ . A routine argument (Andrews and Pollard, 1994; van der Vaart, 1998, Example 19.7, pp. 270–271) yields bracketing numbers of order  $N(\delta, \mathcal{F}) = O(\delta^{-1})$  as  $\delta \rightarrow 0$ ; the bracketing integral is finite, e.g., for  $\gamma = 1$  and  $Q = 2$ . Assumption A(i) can be verified via the bound  $\sup_{\theta \in \Theta} \|f_\theta(\xi_n) - f_\theta(\xi'_n)\|_p \leq \|\xi_n - \xi'_n\|_p$  and (ii) holds trivially.

## Appendix

### A. Proofs

*Proof of the Theorem.* This follows from a simple modification of Andrews and Pollard's (1994) proof of their Theorem 2.1. The proof requires three steps: (i) Their "Proof of inequality (3.2)," (ii) their "Proof of inequality (3.3)," and (iii) their "Comparison of pairs" argument. Replace their  $i$  with  $k$  and their  $\tau(h_i)$  with  $\tau(b_k)$ ; then apply the Lemma below instead of Andrews and Pollard's (1994) Lemma 3.1 in the derivation of their inequality (3.5) to deduce  $\|\max_{1 \leq k \leq N} |\nu_n b_k|\|_Q \leq C' N^{1/Q} \max\{n^{-1/2}, \max_{1 \leq k \leq N} \tau(b_k)\}$  and use this in (i) instead of their inequality (3.5). Another application of the Lemma establishes the required analogue of their inequality (3.5) used in (ii). The same inequality can also be applied in (iii). The other arguments remain valid without changes.  $\square$

**Lemma.** Let  $\tau(f) := \rho(f)^{2/(2+\gamma)}$  for some  $\gamma > 0$  and suppose that Assumption A holds. For all  $n \in \mathbb{N}$ , all  $f, g \in \mathcal{F}$ , and every even integer  $Q \geq 2$  we have

$$\mathbb{E}|\nu_n(f - g)|^Q \leq n^{-Q/2} C \left( (\tau(f - g)^2 n) + \cdots + (\tau(f - g)^2 n)^{Q/2} \right),$$

where  $C$  depends only on  $Q$ ,  $\gamma$ , and  $\alpha$ . The inequality remains valid when  $f - g$  is replaced by  $b_k$  for any given  $k \geq 1$ .

*Proof of the Lemma.* Let  $Z(i) := f(\xi_i) - \mathbb{E}f(\xi_0) - (g(\xi_i) - \mathbb{E}g(\xi_0))$ . Assume without loss of generality that  $|Z(i)| \leq 1$  for all  $i \geq 1$ ; otherwise rescale and redefine  $C$ . Define  $Z'(i) = f(\xi'_i) - \mathbb{E}f(\xi_0) - (g(\xi'_i) - \mathbb{E}g(\xi_0))$  and note that  $\mathbb{E}Z(i) = \mathbb{E}Z'(i) = 0$  for all  $i \in \mathbb{Z}$  and all  $f, g \in \mathcal{F}$  because  $\xi_i$  and  $\xi'_i$  are identically distributed. For fixed  $k \geq 2$ ,  $d \geq 1$ , and  $1 \leq m < k$ , consider integers  $i_1 \leq \cdots \leq i_m \leq i_{m+1} \leq \cdots \leq i_k$  so that  $i_{m+1} - i_m = d$ . Since  $Z(i)$  and  $Z'(i)$  are stationary, repeatedly add and subtract to see that

$$\begin{aligned} & \left| \mathbb{E}Z(i_1)Z(i_2) \cdots Z(i_k) - \mathbb{E}Z(i_1)Z(i_2) \cdots Z(i_m)\mathbb{E}Z(i_{m+1}) \cdots Z(i_k) \right| \\ &= \left| \mathbb{E}Z(i_1 - i_m)Z(i_2 - i_m) \cdots Z(i_k - i_m) \right. \\ & \quad \left. - \mathbb{E}Z(i_1 - i_m)Z(i_2 - i_m) \cdots Z(0)\mathbb{E}Z(d) \cdots Z(i_k - i_m) \right| \\ &\leq \left| \mathbb{E}Z(i_1 - i_m) \cdots Z(0)(Z(d) - Z'(d))Z(i_{m+2} - i_m) \cdots Z(i_k - i_m) \right| \\ & \quad + \sum_{j=2}^{k-m-1} \left| \mathbb{E}Z(i_1 - i_m) \cdots Z(0)Z'(d) \times \cdots \right. \\ & \quad \left. \times (Z(i_{m+j} - i_m) - Z'(i_{m+j} - i_m)) \cdots Z(i_k - i_m) \right| \\ & \quad + \left| \mathbb{E}Z(i_1 - i_m) \cdots Z(0)Z'(d) \cdots Z'(i_k - i_m) \right. \\ & \quad \left. - \mathbb{E}Z(i_1 - i_m) \cdots Z(0)\mathbb{E}Z(d) \cdots Z(i_k - i_m) \right| \end{aligned} \tag{A.1}$$

In particular, the last summand on the right-hand side is zero because  $Z(i_1 - i_m) \cdots Z(0)$  and  $Z'(d) \cdots Z'(i_k - i_m)$  are independent and  $Z(d) \cdots Z(i_k - i_m)$  and  $Z'(d) \cdots Z'(i_k - i_m)$  are identically distributed. For a large enough  $M > 0$  and some  $s > 1$ , Assumption A(i) and distributional equivalence of  $Z(d)$  and  $Z'(d)$  imply  $\|Z(d) - Z'(d)\|_s \leq \|f(\xi_d) - f(\xi'_d)\|_s + \|g(\xi_d) - g(\xi'_d)\|_s \leq 2 \sup_{f \in \mathcal{F}} \|f(\xi_d) - f(\xi'_d)\|_s \leq M\alpha^d$ . Hölder's inequality then bounds the first term on the right-hand side of the preceding display by

$$\|Z(i_1) \cdots Z(i_m)\|_p \|Z(i_{m+2}) \cdots Z(i_k)\|_q M\alpha^d, \quad (\text{A.2})$$

where the reciprocals of  $p$ ,  $q$ , and  $s$  sum to 1. Proceeding similarly to Andrews and Pollard (1994), another application of the Hölder inequality yields

$$\|Z(i_1) \cdots Z(i_m)\|_p \leq \left( \prod_{j=1}^m \mathbb{E}|Z(i_j)|^{mp} \right)^{1/(mp)} \leq \tau(f - g)^{(2+\gamma)/p}$$

whenever  $mp \geq 2$  and similarly  $\|Z(i_{m+2}) \cdots Z(i_k)\|_q \leq \tau(f - g)^{(2+\gamma)/q}$  whenever  $(k - m - 1)q \geq 2$ . Suppose for now that  $k \geq 3$ . If  $k > m + 1$ , take  $s = (\gamma + Q)/\gamma$  and  $mp = (k - m - 1)q = (k - 1)/(1 - 1/s)$ . Decrease the resulting exponent of  $\tau(f - g)$  from  $Q(2 + \gamma)/(Q + \gamma)$  to 2 so (A.2) is bounded by  $M\alpha^d \tau(f - g)^2$ . If  $k \geq 2$  and  $k = m + 1$ , the factor  $\|Z(i_{m+2}) \cdots Z(i_k)\|_q$  is not present in (A.2), but we can still choose  $s = (\gamma + Q)/\gamma$  and  $mp = (k - 1)/(1 - 1/s)$  to obtain the same bound. Identical arguments also apply to each of the other summands in (A.1). Hence, we can find some  $M' > 0$  so that

$$|\mathbb{E}Z(i_1)Z(i_2) \cdots Z(i_k)| \leq |\mathbb{E}Z(i_1)Z(i_2) \cdots Z(i_m)\mathbb{E}Z(i_{m+1}) \cdots Z(i_k)| + M'\alpha^d \tau(f - g)^2.$$

Here  $M'$  in fact depends on  $k$ , but this does not disturb any of the subsequent steps.

Now replace (A.2) in Andrews and Pollard (1994) by the inequality in the preceding display. In particular, replace their  $8\alpha(d)^{1/s}$  with  $M'\alpha^d$  and their  $\tau^2$  with  $\tau(f - g)^2$ . The rest of their arguments now go through without changes. The desired result for  $b_k$  follows mutatis mutandis: Simply define  $Z(i) = b_k(\xi_i)$ , repeat the above steps, and invoke Assumption A(ii) in place of Assumption A(i).  $\square$

## References

- Andrews, D. W. (1994). Empirical process methods in econometrics. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics*, Volume IV, Chapter 37, pp. 2248–2294. Elsevier.
- Andrews, D. W. K. and D. Pollard (1994). An introduction to functional central limit theorems for dependent stochastic processes. *International Statistical Review* 62, 119–132.

- Dedecker, J., P. Doukhan, G. Lang, J. R. León, S. Louhichi, and C. Prieur (2007). *Weak Dependence: With Examples and Applications*. Springer.
- Hagemann, A. (2011). Robust spectral analysis. Unpublished manuscript, Department of Economics, University of Illinois, [arXiv:1111.1965v1](#).
- Hansen, B. E. (1996). Stochastic equicontinuity for unbounded heterogeneous arrays. *Econometric Theory* 12, 347–359.
- Linton, O. and J.-Y. Whang (2007). The quantilogram: With an application to evaluating directional predictability. *Journal of Econometrics* 141, 250–282.
- Pollard, D. (1985). *Convergence of Stochastic Processes*. Springer.
- Shao, X. (2011). Testing for white noise under unknown dependence and its applications to goodness-of-fit for time series models. *Econometric Theory* 27, 312–343.
- Shao, X. and W. B. Wu (2007). Asymptotic spectral theory for nonlinear time series. *Annals of Statistics* 35, 1773–1801.
- van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.
- Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences* 102, 14150–14154.
- Wu, W. B. and W. Min (2005). On linear processes with dependent observations. *Stochastic Processes and their Applications* 115, 939–958.